Ekonomi İşletme ve Maliye Araştırmaları Dergisi
Journal of Economics, Business and Finance Research
e-ISSN: 2667-503X

Yıl: 2024 **Cilt:** 6 **Sayı:** 2 **Sayfa:** 208-220
Makale Geliş Tarihi: 20 Temmuz 2024 **Makale Kabul Tarihi:** 23 Ağustos 2024
Makale Yayın Tarihi: 31 Ağustos 2024 **Doi:** 10.38009/ekimad.1519608

Araştırma Makalesi

Process Capability Analysis of Prediction Data of ML Algorithms

Tuğçe Altuntaş*
Abdulkadir Atalan**

Abstract

This study integrates process capability analysis with Machine Learning (ML) methods to optimize business processes. ML, especially Random Forest (RF) and k-nearest neighbor (kNN) algorithms, has enabled the practical analysis of large data sets by using them together with process capability analysis. This integration enabled real-time monitoring and predictive analytics, enabling the proactive identification of process variations and the making of timely adjustments to maintain or increase process capability. Additionally, ML algorithms have helped optimize process parameters and identify critical factors affecting process performance, allowing for continuous improvement and achieving desired quality standards with greater efficiency. In conclusion, this study provides the basis for the synergy between process capability analysis and ML methods to enable businesses to achieve higher levels of quality control, productivity, and competitiveness in dynamic and complex production environments.

Keywords: *Process Capability Analysis, Machine Learning, Random Forest, k-nearest Neighbor*

Jel Classification: *C15, C44, C46*

* Doctorate, Çanakkale Onsekiz Mart University, Faculty of Policical Science, Department of International Business, tugcealtuntas@comu.edu.tr ORCID NO: 0009-0007-3629-2454

** Assoc. Prof. Dr., Çanakkale Onsekiz Mart Unıversıty, Faculty of Engıneerıng, Department of Industrıal Engıneerıng, abdulkadiratalan@gmail.com ORCID NO: 0000-0003-0924-3685

ML Algoritmalarının Tahmin Verilerine ait Süreç Yeteneği Analizi

Öz

Bu çalışma, iş süreçlerini optimize etmek için process yetenek analizini Makine Öğrenimi (ML) yöntemleriyle entegre etmekteyi amaçlamıştır. ML, özellikle Rastgele Orman (RF) ve k-en yakın komşu (kNN) algoritmaları, süreç yetenek analizi ile birlikte kullanılarak büyük veri setlerinin pratik analizine olanak sağlamıştır. Bu entegrasyon, gerçek zamanlı izleme ve tahmine dayalı analitiği mümkün kılarak süreç değişimlerinin proaktif olarak belirlenmesine ve süreç yeterliliğini korumak veya artırmak için zamanında ayarlamalar yapılmasına olanak sağladı. Ek olarak, ML algoritmaları süreç parametrelerinin optimize edilmesine ve süreç performansını etkileyen kritik faktörlerin belirlenmesine yardımcı olarak sürekli iyileştirmeye ve istenen kalite standartlarına daha yüksek verimlilikle ulaşılmasına olanak tanıdı. Sonuç olarak bu çalışma, işletmelerin dinamik ve karmaşık üretim ortamlarında daha yüksek düzeyde kalite kontrol, üretkenlik ve rekabet gücü elde etmelerini sağlamak için süreç yetenek analizi ile ML yöntemleri arasındaki sinerjinin temelini oluşturmaktadır.

Anahtar Kelimeler: *Süreç Yeteneği Analizi, Makine Öğrenimi, Rastgele Orman, k-En Yakın Komşu*

Jel Sınıflandırması: *C15, C44, C46*

1. Introduction

Process capability analyses are a critical tool used to evaluate the functions and operations of an organization or business (Van Looy, 2020). These analyses are often used to develop strategies to increase business processes' effectiveness, efficiency, and performance (Y. A. Atalan & Atalan, 2023). Competency analysis helps determine the appropriate business processes and areas needing improvement to help the organization achieve its goals (Campion et al., 2011).

Different methods are used to analyze the process capability of a business or organization (Chen et al., 2014). These include techniques such as workflow analysis, performance metrics and performance indicators, data analysis, customer feedback, and employee testimonials (Ayaz Atalan & Atalan, 2020). These methods are combined to assess the organization's current state and identify its potential for improvement. The primary purpose of process capability analysis is to increase the effectiveness of processes to increase the business's competitive advantage (Ray et al., 2004). These analyses help develop strategies to identify and eliminate obstacles to achieving the business's goals. In addition, process capability analyses also help reduce operational costs to ensure more efficient use of resources and increase the business's profitability (Migdadi, 2022; Ray et al., 2004).

Process capability analyses are essential for evaluating and improving an organization's business processes (Kerpedzhiev et al., 2021). Process capability analyses help businesses develop strategies to increase competitive advantage, reduce operational costs, and increase customer satisfaction (A. Atalan, 2020; Puspita et al., 2020). Organizations can operate more effectively and efficiently and gain a competitive advantage based on the results of process capability analyses (A. Atalan, 2022; Gupta et al., 2020). The relationship between process capability analysis and prediction data based on ML algorithms creates a solid relationship to increase the operational efficiency of businesses and gain competitive advantage (A. Atalan, 2023). Process capability analyses provide a comprehensive approach to evaluating the effectiveness and efficiency of existing business processes (Kerpedzhiev et al., 2021). These analyses identify the business's operational weak points and improvement potential, while ML algorithms help predict future trends and possible scenarios by analyzing large amounts of data.

ML algorithms help businesses make operational decisions using data obtained from process capability analyses (Bharadiya, 2023). In particular, process capability analysis in business supply chain management evaluates supplier performance, while ML algorithms have provided positive advantages in predicting future supply demands based on historical data (Cavalcante et al., 2019). Thus, it allows businesses to optimize supply chain costs and increase efficiency. Likewise, process adequacy analysis in areas such as customer relationship management evaluates the service processes of companies to improve customer satisfaction (Reinartz et al., 2004). At the same time, ML algorithms pave the way for predicting future demands by analyzing customer behavior. This gives businesses essential data to make strategic decisions to increase customer satisfaction and expand market share.

ML covers a broad field of algorithms that extract meaningful information from data and make predictions (A. Atalan et al., 2022). These algorithms are divided into supervised, unsupervised, and reinforcement learning. Supervised learning algorithms train using labeled data sets and learn to predict outputs corresponding to specific inputs (Talukdar & Biswas, 2024). Examples of this category include algorithms such as linear regression, support vector machines (SVM), decision trees, k-nearest neighbor (kNN), and random forest (RF). Unsupervised learning algorithms aim to discover hidden structures in data by working on unlabeled data (Naeem et al., 2023). Clustering algorithms (e.g., k-means) and dimensionality reduction techniques fall into this group. Reinforcement learning, on the other hand, refers to learning based on the dynamics between an agent and the environment in which it interacts and is often used in game theory and robotics (Abouelyazid, 2024). The choice of algorithm varies depending on the type of problem, size of the data set, type of data, and desired results (Ayaz Atalan et al., 2020).

Random Forest (RF) and k-nearest Neighbor (kNN) are popular methods among supervised learning algorithms (İnaç et al., 2022). RF is an ensemble method created by training multiple decision trees on random subsets (Thakur & Biswas, 2024). Each tree is trained on different subsets of the data set and randomly selected subsets of its features. This diversity increases the model's generalization ability and reduces the risk of overlearning. RF can be used in classification and regression problems and offers high accuracy rates. The RF model can also be used to determine the importance of variables, increasing the model's interpretability. The computational cost of RF is higher than that of a single decision tree, but these costs can be reduced with parallel processing (Ferro et al., 2023).

The k-nearest Neighbor (kNN) algorithm is another supervised learning method known for its simplicity and intuitive nature (Sabry, 2023). kNN looks at the k nearest neighbors in the data set to determine the class or value of a new data point (Ukey et al., 2023). This algorithm does not require model training and stores data directly. When making the prediction, the distances between the new data point and the entire training data set are calculated, and a prediction is made using the class or value of the k nearest neighbors. The performance of kNN depends on the value of k and the distance metric used (Inyang et al., 2023). Although it is effective on small data sets, computational cost may increase, and performance may decrease in large and high-dimensional data sets. For this reason, kNN is often used for early-stage analyses or to benchmark the performance of other, more complex models (Prasad et al., 2023).

The use of RF and kNN algorithms in scientific studies is quite common due to the advantages provided by these algorithms (Zhang et al., 2023). RF delivers reliable results even on complex data sets, with high accuracy rates and the ability to reduce the risk of overlearning. Additionally, determining the importance of variables helps researchers understand essential features in the data set. This feature is precious in bioinformatics, medicine, finance, and social sciences (A. Atalan & Atalan, 2022). kNN, however, is an ideal option for quick initial analyses, especially with its simple and understandable structure (Chafai et al., 2024). Additionally, its direct applicability on labeled datasets and its non-parametric nature provide flexibility across many data types. Using these two algorithms together allows for analyzing data from different perspectives and comparing the results (Khan et al., 2024). Therefore, RF and kNN were preferred in this study as they are practical and complementary tools for various data types and problems (İnaç et al., 2022).

As a result, the relationship between process capability analyses and ML algorithms requires strong collaboration to increase the operational efficiency of businesses and gain competitive advantage.

Data from process capability analysis supports the predictive capabilities of ML algorithms and provides businesses with a solid foundation to support future decisions. This allows companies to gain a competitive advantage and drive continuous improvement.

This study consists of four main sections. The first section includes general information about the study topic, the method used, and a literature review. The second section of the study consists of a statistical definition of the data used for the study, the Process capability analysis method, and ML algorithms selected as methods. The third section of the study discusses the results obtained as a result of the data analysis in the method section of the study with the techniques used. The final section of the study includes the conclusion section.

2. Methodology

This study created data sets of three independent and dependent variables to verify the validity of the prediction results obtained in ML algorithms within the scope of process capability analysis for a business. Descriptive statistics data for dependent and independent variables are presented in **Table 1.** Within the scope of the research, prediction models were developed using Random Forest (RF) and k-nearest Neighbor (kNN) algorithms, and the performances of these models were evaluated. The methodology section explains the preparation of data sets, the structuring of the algorithms, and the model evaluation criteria in detail. This approach aims to determine the effectiveness and accuracy of machine learning algorithms in process capability analysis.

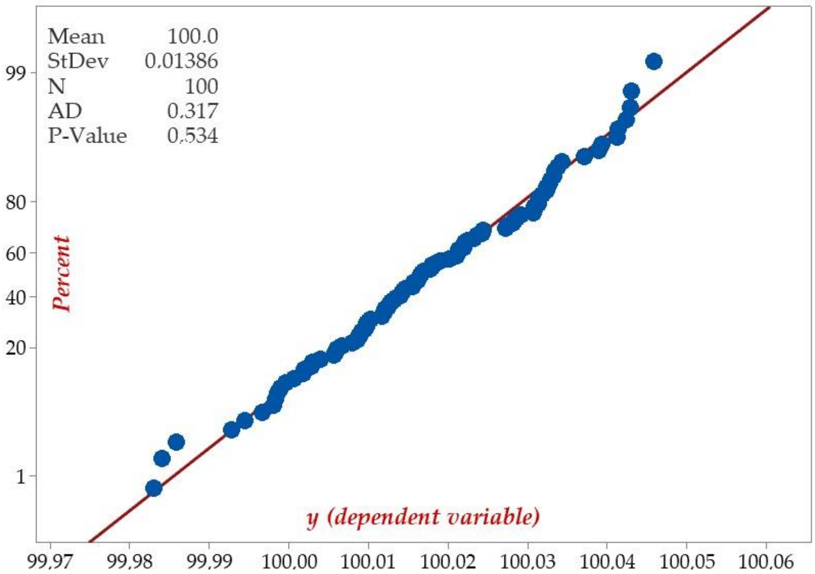
Table 1: Descriptive statistics data for dependent and independent variables

Variable	x_1	x_2	x_3	y
Total Sample Size	100	100	100	100
Mean	0.20	20.98	2.65	100.02
Standard Error of Mean	0.00	0.43	0.14	0.00
Standard Deviation	0.02	4.31	1.35	0.01
Variance	0.00	18.56	1.83	0.00
Coefficient of Variance	8.98	20.54	50.99	0.01
Minimum	0.16	11.33	0.00	99.98
Q1 (The first quartile)	0.19	17.67	2.00	100.01
Median	0.20	21.07	3.00	100.02
Q3(The third quartile)	0.21	24.13	4.00	100.03
Maximum	0.25	29.06	6.00	100.05
Range	0.09	17.73	6.00	0.06
Inter Quartile Range	0.03	6.46	2.00	0.02
N for Mode	0.00	0.00	28.00	0.00
Skewness	0.12	-0.10	0.19	-0.15
Kurtosis	-0.52	-0.66	-0.50	-0.31

The mean values of the variables x_1, x_2, and x_3 are 0.20, 20.98, and 2.65, respectively. The standard deviations of variables are 0.02, 4.31, and 1.35, respectively. These statistics show that the variable x_2 has a more considerable variance and dispersion than the others. On the other hand, the mean value for the variable y was determined to be 100.02, and the standard deviation was 0.01, indicating that the distribution has a relatively low variance and a low coefficient.

Regarding quartiles and median values, there appears to be no significant difference between quartiles for variables x_1 and x_3, indicating that the distributions are symmetric. However, a more substantial difference between quartiles for the x_2 variable suggests that the distribution could be more balanced. Skewness and kurtosis values determine the symmetry and flatness/steepness properties of the distributions of the variables. These values show that variables x_1 and x_3 are symmetrical, while variable x_2 shows a slight negative skew. The skewness and kurtosis values for the y variable are low, indicating that the distribution is more symmetrical and normal. The normal distribution graph of the data of the dependent variable is shown in **Figure 1.**

Figure 1: The normal distribution graph of the data of the dependent variable



This study used process adequacy analysis to verify the validity of the prediction results obtained in ML algorithms. Among many ML models, only RF and kNN algorithms were preferred in this study. RF is a robust ensemble learning algorithm frequently used in data mining and ML. This algorithm works by creating and combining multiple decision trees. Each decision tree is trained using a randomly selected subset of the data set and a random subset of features. This process increases the model's generalization ability and significantly reduces the risk of overfitting. RF can be used in both classification and regression problems and has the potential to achieve high accuracy rates. This method can also be used to determine the importance of variables in the data set, which is useful for feature selection and model interpretation. Advantages of RF include the ability to work with various data types, deal with missing data, and process large data sets. However, training the model and making predictions may require more computing than a single decision tree.

The kNN algorithm is a classification and regression method in the supervised learning category known primarily for its simplicity and effectiveness. kNN looks at the class or value of its k nearest neighbors in the data set to determine the class or value of a new data point. This algorithm is based on the principle that similar things are found in similar places. The main advantage of kNN is that it does not require model training; the data is stored directly and processed during querying. This is especially ideal for small and medium-sized datasets. However, in large data sets, computational costs may increase, and in high-dimensional data sets, difficulties may arise, such as the concept of proximity losing meaning. Additionally, the choice of k value and distance metric directly affects the model's performance, and these choices may vary from problem to problem. kNN is often used for early-stage analysis or to benchmark the performance of other more complex models. Prediction data were obtained by running ML algorithms in the open-access Orange Data Mining software. The process flow chart of ML algorithms is visualized in **Figure 2**.

Figure 2: The process flow chart of ML algorithms

Hyperparameter information of RF and kNN algorithms is given in **Table 2 and Table 3**. The RF model consists of 10 trees, each of which is a different decision tree. There is no limit on the maximum number of features, meaning all features are evaluated at each node. The model training is structured to produce different results in each run and is not repeatable. The depth of the trees is unlimited, with no maximum depth limit. Nodes stop splitting when they contain a maximum of 5 samples. The training dataset consists of 80 data samples with three features x_1, x_2, x_3. The target variable is y. This configuration allows the model to build flexible and potentially deep trees, producing different results in each run and working on small datasets.

Table 2: The parameters of the RF algorithm

Number of trees	10
Maximal number of considered features	Unlimited
Replicable training	No
Maximal tree depth	Unlimited
Stop splitting nodes with maximum instances	5
Data instances	80
Features	x_1, x_2, x_3
Target	y

Table 3: The parameters of the kNN algorithm

Number of neighbors	5
Metric	Euclidean
Weight	Uniform
Data instances	80
Features	x_1, x_2, x_3
Target	y

The parameters of the kNN algorithm describe how the model is structured and run. This model considers 5 neighbors and uses the Euclidean metric to calculate the distance between neighbors. The weights of neighboring samples are equal, and each has the same effect on the model (uniform weight). The training dataset consists of 80 data samples with three features x_1, x_2, x_3. The target variable is y. This configuration evaluates the distances of the samples by determining the 5 nearest neighbors for each sample and weighting them equally when performing classification or regression.

Process capability analysis is a technique used to evaluate the effectiveness and efficiency of a specific process of a business or organization. This analysis measures a process's ability, performance, and suitability to achieve specific objectives. Process capability analysis generally evaluates the current state of a particular process, determines its potential to achieve the business's goals, and develops strategies for process improvement.

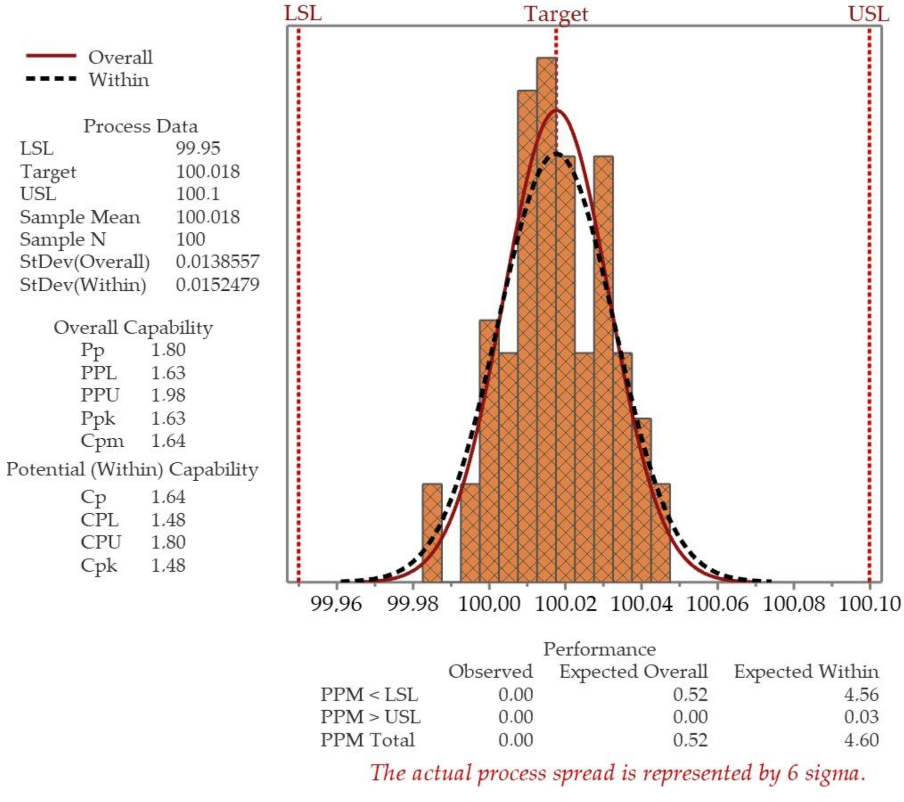
Various statistical measurements, such as the Cp value, are often used during process capability analysis. Cp is an index that measures a process's accuracy rate and continuity. The Cp value shows how well the process complies with production tolerances. The Cp value is used to classify situations where the process operates correctly into appropriate ranges. The Cp value shows how well a process complies with production tolerances. This value indicates how stable and low-variance the processes are.

In general, the Cp value of a process should be greater than 1.00. A Cp value greater than 1.00 indicates that the process operates stably within the tolerance range and exhibits slight variance. The higher the Cp value, the higher the process's compliance with production tolerances. However, the lower the Cp value, the less ability the process has to operate within the tolerance range, and there is room for improvement. Ideally, the Cp value should be 1.33 or higher. This indicates that the process is operating stably and within production tolerances. In this study, the tolerance properties of the process were calculated by calculating the CP values of the prediction results obtained from ML algorithms.

3. Results

In this study, process adequacy indices of a system were calculated by deriving 100 data regarding dependent and independent variables. The process adequacy index data and graph of the sample data are shown in **Figure 3.**

Figure 3: Process capability index data and graph of sample data



The process's lower and upper limits (LSL and USL) were determined first, and the target value (Target) was defined as 100.018. These values are used to determine the acceptable production range of the process. The average value (Sample Mean) obtained from the sampling was 100.018, which is very close to the target value. The total number of samples (Sample N) is also reported as 100. These data show that the process works quite close to the desired goal. Considering the standard deviations, the overall standard deviation (StDev(Overall)) was calculated as 0.0138557, and the within-sample standard deviation (StDev(Within)) was calculated as 0.0152479. These values are essential to evaluate the internal variability and stability of the process.

The values calculated for Overall Capability determine the production capacity of the process and its proximity to the target. The Pp (Process Potential) value was reported as 1.80, showing that the process has a wide distribution within the acceptable production range. PPL (Process Potential Lower) and PPU (Process Potential Upper) are determined as 1.63 and 1.98, respectively; these values show the potential width of the process according to its lower and upper limits. Ppk (Process Potential Capability Index) and Cpm (Process Capability Index with Overall Standard Deviation) evaluate the closeness and variability of the process to the target.

The calculated values for potential (in-sample) competence are also crucial in evaluating the performance of the process. The Cp (Process Capability Index) value is determined as 1.64, showing how well the process works within the desired tolerance range. CPL (Process Capability Lower) and CPU (Process Capability Upper) determine the capacity of the process according to its lower and upper limits. Finally, the Cpk (Process Capability Index with Process Average) value is reported as 1.48 and shows the closeness and variability of the process to the target. These values show how reliably the process operates within the desired tolerance range.

Additionally, observed errors (Observed) and expected errors (Expected) in a particular process were compared for performance evaluation. First, it can be seen that values lower than the lower limit (LSL) (PPM < LSL) are not observed at all. It is reported that such expected errors are 0.52. This shows that the process does not produce defective products below the lower limit and is at the desired level in terms of performance. It was stated that values higher than the upper limit (USL) (PPM > USL) were not observed. It is reported that such expected errors are only 0.03. This shows that the process stays within the upper limit, and the products are generally within the upper limits. When total errors (PPM Total) were examined, it was stated that no errors were observed in total. The expected total errors were reported to be 4.60. These results show that the process is stable and produces the desired results within targeted limits. **Table 4** includes the performance criterion data of the RF and kNN algorithms.

Table 4: The performance criterion data of the RF and kNN algorithms

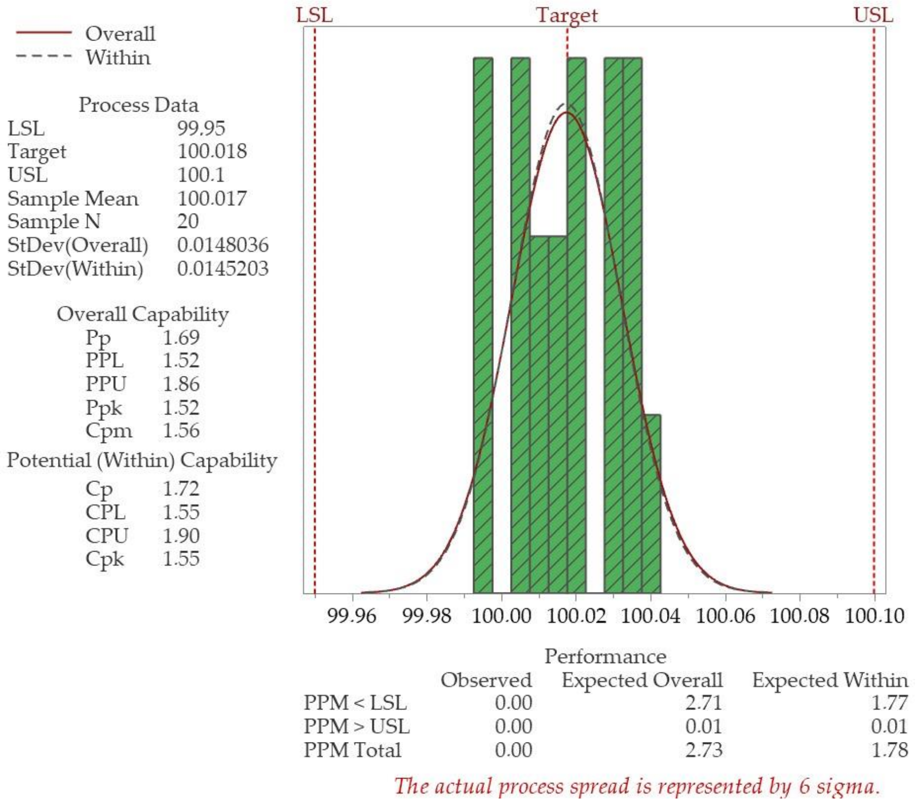
Model	MSE	RMSE	MAE	MAPE	R^2
RF	0.001	0.004	0.003	0.001	0.925
kNN	0.001	0.007	0.005	0.001	0.806

We see that both algorithms have similar mean square error (MSE) values: 0.001 for RF and 0.001 for kNN when we compare the performance benchmarks of the RF and kNN algorithms. However, the kNN algorithm has higher root mean square error (RMSE) and mean absolute error (MAE) values than RF: 0.007 and 0.005, respectively, versus 0.004 and 0.003, respectively, for RF. The predictions of the RF algorithm have a lower error level than kNN. Similarly, the mean absolute percent error (MAPE) value for the kNN algorithm is slightly higher than that for RF (0.001 versus 0.001), indicating that RF's predictions are more consistent. Moreover, the determined R-squared (R^2) values also show that the RF algorithm has a higher explanatory power: 0.925 versus 0.806 of kNN. It is understood that the RF algorithm adapts better to the data and makes better predictions. However, both algorithms' relatively low error rates and high R-squared values indicate that they fit the data well and make reasonable predictions.

Figure 4 shows the results of the process capability analysis of the dependent variable data set predicted by an RF (Random Forest) algorithm. First, the lower and upper limits of the process were determined, and the target value was defined as 100.018. The average value obtained during the sampling process was reported as 100,017, and the total number of samples was 20. The overall standard deviation was calculated as 0.0148036, and the within-sample standard deviation was

0.0145203. These data show that the process works close to the desired target, and standard deviations are acceptable.

Figure 4: Process capability index data and graph of RF algorithm

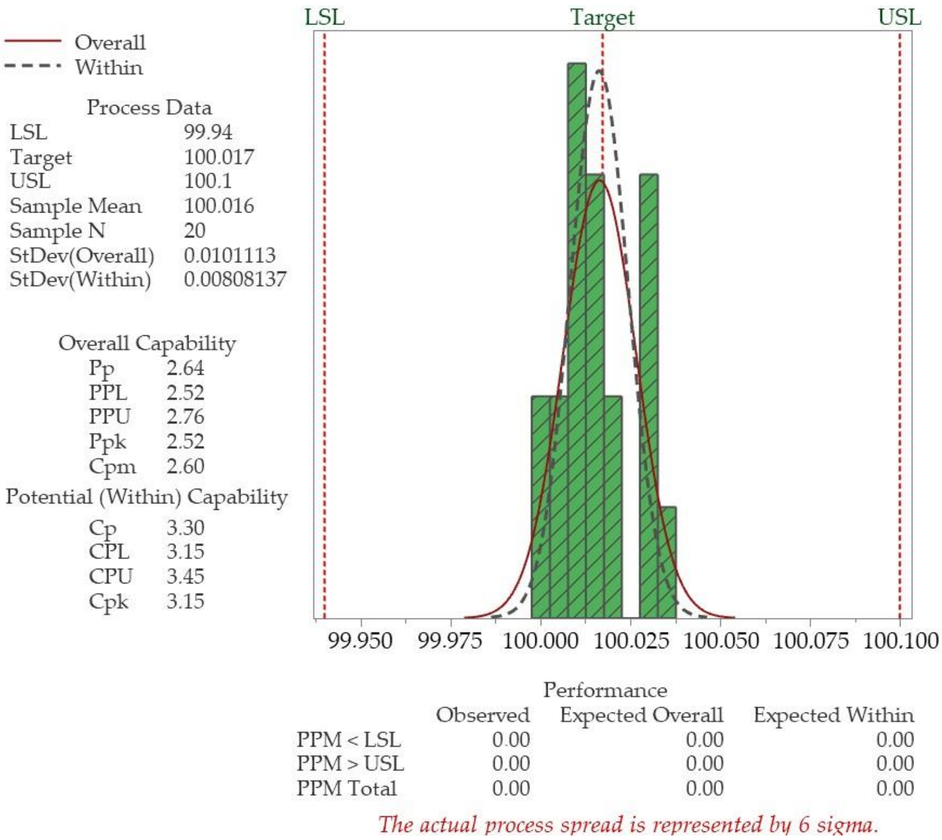


The Pp value is reported as 1.69 and determines the production capacity of the process when the calculated values for general competence are examined. PPL and PPU are defined as 1.52 and 1.86, respectively, and show the potential width of the process according to its lower and upper limits. Ppk and Cpm evaluate the closeness and variability of the process to the target. The calculated values for potential (in-sample) competence are also crucial in assessing the performance of the process. The Cp value was determined to be 1.72, showing how well the process works within the desired tolerance range. CPL and CPU determine the capacity of the process according to its lower and upper limits. Finally, the Cpk value was reported as 1.55, indicating the closeness and variability of the process to the target.

It was stated that values lower than the lower limit (PPM < LSL) were not observed when the performance evaluation was examined. The expected errors were reported to be 2.71. It was stated that values higher than the upper limit (PPM > USL) were not observed, and such expected errors were only 0.01. When the total errors (PPM Total) were examined, it was reported that no errors were observed, and the expected total errors were 2.73. These results show that the process is stable and produces the desired results within targeted limits.

Figure 5 presents the results of the process capability analysis of the dependent variable data set predicted by a kNN algorithm. First, the lower and upper limits of the process were determined, and the target value was defined as 100.017. The average value obtained during the sampling process was reported as 100,016, and the total number of samples was 20. The overall standard deviation was calculated as 0.0101113, and the within-sample standard deviation was calculated as 0.00808137. These data show that the process works close to the desired target, and the standard deviations are pretty low.

Figure 5: Process capability index data and graph of the kNN algorithm



In terms of the values calculated for overall capability, the Pp value is reported as 2.64 and determines the production capacity of the process. PPL and PPU are defined as 2.52 and 2.76, respectively, and indicate the potential width of the process relative to its lower and upper limits. Ppk and Cpm evaluate the closeness and variability of the process to the target. The calculated values for potential (in-sample) competence are also crucial for assessing the performance of the process. The Cp value was determined to be 3.30, showing how well the process works within the desired tolerance range. CPL and CPU determine the capacity of the process according to its lower and upper limits. Finally, the Cpk value was reported as 3.15, indicating the closeness and variability of the process to the target.

In terms of performance evaluation, it is stated that values lower than the lower limit (LSL) (PPM < LSL) or higher than the upper limit (USL) (PPM > USL) are not observed. This shows that the process operates flawlessly within the desired tolerance range. When the total errors (PPM Total) were examined, it was stated that no errors were observed, and the expected total errors were reported to be 0.00. These results show that the process is perfectly stable and produces the desired results within targeted limits.

As a result, the process adequacy indexes of the prediction data obtained by ML algorithms are calculated and contribute to the correct management of the performance of the processes for the future. For this reason, integrating process capability analysis and ML methods confirms with this study that a company has the right processes for future periods.

Integrating process capability analysis with methods offers a robust approach to enhance quality management and optimize processes. By leveraging ML algorithms, organizations can analyze large volumes of data more efficiently and accurately, identifying patterns and insights that may not be apparent through traditional statistical methods alone. This integration enables real-time monitoring and predictive analytics, proactively identifying process variations or abnormalities, thus facilitating timely adjustments to maintain or improve process capability. Additionally, it can assist in optimizing process parameters and identifying key factors that influence process performance, leading to continuous improvement and greater efficiency in achieving desired quality standards. Overall, the combination of process capability analysis and methods empowers organizations to achieve higher levels of quality control, productivity, and competitiveness in dynamic and complex manufacturing environments.

4. Conclusion

This study shows that combining process capability analysis with ML algorithms helps businesses gain a competitive advantage by increasing their operational efficiency. Process capability analysis allows companies to evaluate their current processes regarding effectiveness and efficiency. At the same time, ML algorithms help them predict future trends and possible scenarios by analyzing large data sets. This integration enables real-time monitoring and predictive analytics, enabling businesses to identify process variations promptly and make necessary adjustments to maintain or increase process capability. Additionally, ML can help optimize process parameters and identify critical factors affecting process performance, enabling continuous improvement and greater efficiency in achieving desired quality standards. Ultimately, the combination of process capability analysis and methods allows organizations to achieve higher levels of quality control, productivity, and competitiveness in dynamic and complex manufacturing environments.

CONTRIBUTION OF AUTHORS

The authors contributed equally to the creation of this article.

CONFLICT OF INTEREST DECLARATION

There is no financial conflict of interest with any institution, organization, or person, and there is no conflict of interest among the authors.

REFERENCES

Abouelyazid, M. (2024). Reinforcement Learning-based Approaches for Improving Safety and Trust in Robot-to-Robot and Human-Robot Interaction. *Advances in Urban Resilience and Sustainable City Design*, *16*(02), 18–29.

Atalan, A. (2020). Logistic Performance Index of OECD Members. *Akademik Araştırmalar ve Çalışmalar Dergisi*, *12*(23), 608–619. https://doi.org/10.20990/kilisiibfakademik.720604

Atalan, A. (2022). Desirability Optimization Based on the Poisson Regression Model: Estimation of the Optimum Dental Workforce Planning. *International Journal of Health Management and Tourism*, *7*(2), 200–216. https://doi.org/10.31201/ijhmt.1123824

Atalan, A. (2023). Integration of Discrete-Event Simulation and Statistical Process Control Methods. *1st International Conference on Pioneer and Innovative Studies*, *1*(1), 38–46.

Atalan, A., & Atalan, Y. A. (2022). Analysis of the Impact of Air Transportation on the Spread of the COVID-19 Pandemic. In G. Catenazzo (Ed.), *Challenges and Opportunities for Transportation Services in the Post-COVID-19 Era* (pp. 68–87). IGI Global. https://doi.org/10.4018/978-1-7998-8840-6.ch004

Atalan, A., Şahin, H., & Atalan, Y. A. (2022). Integration of Machine Learning Algorithms and Discrete-Event Simulation for the Cost of Healthcare Resources. *Healthcare*, *10*(10), 1920. https://doi.org/10.3390/healthcare10101920

Atalan, Y. A., & Atalan, A. (2023). Integration of the Machine Learning Algorithms and I-MR Statistical Process Control for Solar Energy. *Sustainability*, *15*(18), 13782. https://doi.org/10.3390/su151813782

Ayaz Atalan, Y., & Atalan, A. (2020). A Statistical Analysis of the Relationship Between Meteorological Parameters and the Spread of COVID-19 Cases: Comparison Between Turkey and Italy. *Journal of Statistics and Applied Sciences*, *1*(2), 76–84.

Ayaz Atalan, Y., Tayanç, M., Erkan, K., & Atalan, A. (2020). Development of Nonlinear Optimization Models for Wind Power Plants Using Box-Behnken Design of Experiment: A Case Study for Turkey. *Sustainability*, *12*(15), 6017. https://doi.org/10.3390/su12156017

Bharadiya, J. P. (2023). The role of machine learning in transforming business intelligence. *International Journal of Computing and Artificial Intelligence*, *4*(1), 16–24.

Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology*, *64*(1), 225–262.

Cavalcante, I. M., Frazzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. *International Journal of Information Management*, *49*, 86–97.

Chafai, N., Bonizzi, L., Botti, S., & Badaoui, B. (2024). Emerging applications of machine learning in genomic medicine and healthcare. *Critical Reviews in Clinical Laboratory Sciences*, *61*(2), 140–163.

Chen, Y., Wang, Y., Nevo, S., Jin, J., Wang, L., & Chow, W. S. (2014). IT capability and organizational performance: the roles of business process agility and environmental factors. *European Journal of Information Systems*, *23*(3), 326–342.

Ferro, M., Silva, G. D., de Paula, F. B., Vieira, V., & Schulze, B. (2023). Towards a sustainable artificial intelligence: A case study of energy efficiency in decision tree algorithms. *Concurrency and Computation: Practice and Experience*, *35*(17), e6815.

Gupta, S., Drave, V. A., Dwivedi, Y. K., Baabdullah, A. M., & Ismagilova, E. (2020). Achieving superior organizational performance via big data predictive analytics: A dynamic capability view. *Industrial Marketing Management*, *90*, 581–592.

İnaç, H., Ayözen, Y. E., Atalan, A., & Dönmez, C. Ç. (2022). Estimation of Postal Service Delivery Time and Energy Cost with E-Scooter by Machine Learning Algorithms. *Applied Sciences*, *12*(23), 12266. https://doi.org/10.3390/app122312266

Inyang, U. G., Ijebu, F. F., Osang, F. B., Afolorunso, A. A., Udoh, S. S., & Eyoh, I. J. (2023). A Dataset-Driven Parameter Tuning Approach for Enhanced K-Nearest Neighbour Algorithm Performance. *International Journal on Advanced Science, Engineering & Information Technology*, *13*(1).

Kerpedzhiev, G. D., König, U. M., Röglinger, M., & Rosemann, M. (2021). An exploration into future business process management capabilities in view of digitalization: results from a Delphi study. *Business & Information Systems Engineering*, *63*(2), 83–96.

Khan, N., Mohmand, M. I., Rehman, S. ur, Ullah, Z., Khan, Z., & Boulila, W. (2024). Advancements in intrusion detection: A lightweight hybrid RNN-RF model. *Plos One*, *19*(6), e0299666.

Migdadi, M. M. (2022). Knowledge management processes, innovation capability and organizational performance. *International Journal of Productivity and Performance Management*, *71*(1), 182–210.

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*.

Prasad, B. V. V. S., Gupta, S., Borah, N., Dineshkumar, R., Lautre, H. K., & Mouleswararao, B. (2023). Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach. *Preventive Medicine*, *174*, 107619.

Puspita, L. E., Christiananta, B., & Ellitan, L. (2020). The effect of strategic orientation, supply chain capability, innovation capability on competitive advantage and performance of furniture retails. *International Journal of Scientific & Technology Research*, *9*(03), 4521–4529.

Ray, G., Barney, J. B., & Muhanna, W. A. (2004). Capabilities, business processes, and competitive advantage: choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal*, *25*(1), 23–37.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research*, *41*(3), 293–305.

Sabry, F. (2023). *K Nearest Neighbor Algorithm: Fundamentals and Applications* (Vol. 28). One Billion Knowledgeable.

Talukdar, W., & Biswas, A. (2024). Synergizing Unsupervised and Supervised Learning: A Hybrid Approach for Accurate Natural Language Task Modeling. *ArXiv Preprint ArXiv:2406.01096*.

Thakur, D., & Biswas, S. (2024). Permutation importance based modified guided regularized random forest in human activity recognition with smartphone. *Engineering Applications of Artificial Intelligence*, *129*, 107681. https://doi.org/https://doi.org/10.1016/j.engappai.2023.107681

Ukey, N., Yang, Z., Li, B., Zhang, G., Hu, Y., & Zhang, W. (2023). Survey on exact knn queries over high-dimensional data space. *Sensors*, *23*(2), 629.

Van Looy, A. (2020). Capabilities for managing business processes: a measurement instrument. *Business Process Management Journal*, *26*(1), 287–311.

Zhang, C., Liu, Y., & Tie, N. (2023). Forest Land Resource Information Acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Trees and Multi-Layer Perceptron. *Forests*, *14*(2), 254.